~~File Copy~~

NATIONAL BUREAU OF STANDARDS REPORT

14 March 1952

NBS Project
1103-21-5106

NBS Report
No. 1521

PROGRESS REPORT FOR JANUARY-MARCH 1952

on

APPLICATION OF STATISTICAL THEORY OF EXTREME VALUES TO GUST-LOAD PROBLEMS

(NACA Project W4712; NBS Project 1103-21-5106)

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

PROGRESS REPORT FOR JANUARY-MARCH 1952

on

APPLICATION OF STATISTICAL THEORY OF EXTREME VALUES TO GUST-LOAD PROBLEMS

(NACA Project W4712; NBS Project 1103-21-5106)

I. SUMMARY

Work during the present quarter (January-March 1952) has been marked by a discovery that radically alters the basic computation techniques for the above project. The new method involves a simple means of expressing the covariances of the order statistics in a random sample of n observations from the extreme-value distribution in terms of tabulated functions. As a result, the scope of the results to be expected can be greatly extended without much, if any, increase in cost over what had been planned.

The present report gives a brief summary of what has been accomplished and what can be expected under the new techniques, and includes tentative plans for preparing the results for publication. This report presents only the main lines of development and avoids most of the mathematical details, since it is intended to serve as a basis for discussion of further work by representatives of the National Bureau of Standards and the National Advisory Committee for Aeronautics at a meeting to be arranged in the near future. The balance of work under this project, as outlined herein, is to be planned with the goal of completion by the end of FY 1952.

II. DESCRIPTION OF RECENT WORK

A. Background.

The specific objective is to obtain a statistical function which will provide improved methods of estimating maximum values of acceleration increments and gust velocities which may be expected by an airplane in flight.

By maximum acceleration increment is meant the largest value occurring during a single flight of a given airplane. If a series of n flights of the same plane are considered, then there will be a maximum value X for each and the set of the n maxima, $X_1, X_2, \ldots, X_n$, constitutes a sample of n observations to be analyzed. One objective of the analysis is to predict a value such that one may, in a long series of flights, expect that the proportion P of the flight maxima will not exceed this value, while the remaining proportion, $1 - P$, will exceed (or equal) it. This upper limit is designated as X_P and naturally depends upon P. For example, if we want to estimate a limit X_P such that in only a very small proportion $(1 - P)$ of the flights will a larger value occur, then X_P must be expected to be quite large. The limit X_P is not known but must be estimated from sample data such as the set of n maxima $X_1, X_2, \ldots, X_n$ mentioned above.

The method of estimation is to find a function $f = f(X_1, X_2, \ldots, X_n)$ of these n variables, called an estimator of X_P, which conforms to the following two desirable characteristics as closely as possible:

(1) The estimator is unbiased: $E[f(X_1, X_2, \ldots, X_n)] = X_P$, where E denotes mathematical expectation. This means that the estimator f fluctuates about a long-run average which is the correct value, X_P.

(2) The estimator is most efficient, that is, has minimum variance: $\sigma^2(f) = E(f-X_P)^2$, where σ^2 denotes variance. This means that the values taken by the estimator f are concentrated so closely about the desired true value X_P that of all unbiased estimators of X_P it has minimum mean squared error.

It is clear that the values of a function f possessing the above two properties may be expected to give very satisfactory estimates of the unknown value X_P. However, in order to apply these two properties, some assumption must be made about the form of the statistical population from which the observations are assumed to come. The fact that each observation is itself a maximum of many individual values encountered in an individual flight, together with other supporting data discussed by Press (reference 1), gives theoretical ground for assuming the underlying population to be of the extreme-value type studied by Dr. E. J. Gumbel, namely $F(x) = \exp(-e^{-y})$, $y = \alpha(x-u)$, where $F(x)$ denotes the cumulative distribution function.

Methods of estimation in present use by the NACA involve the mean and standard deviation of a sample of given data. These methods have been developed principally by Dr. Gumbel. They are not very complicated, but have the disadvantage that their bias and efficiency have not been evaluated because of the great amount of calculation that would be necessary in order to determine the expected values and variances of the functions involved in Dr. Gumbel's estimators.

B. Current progress.

Research to date under the present project has concentrated on building up a simple type of estimator f which is a linear function of the order statistics of the sample. That is, if we arrange the n observations in increasing order of size and let x_1 denote the smallest and x_n the largest, then the sample may be represented by

$$(x_1, x_2, \ldots, x_n), \quad x_1 \leq x_2 \leq \ldots \leq x_n ,$$

where the x_i are called order statistics of the sample, and the linear estimator sought is of the form

$$T_n = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n .$$

The weights w_i are to be determined so as to satisfy conditions (1) and (2) above, namely:

(1) T_n is unbiased: $\sum_{i=1}^{n} w_i E x_i = X_p$;

(2) T_n has minimum variance: $\sigma^2(T) = \sum_{j=1}^{n} \sum_{i=1}^{n} \sigma_{ij} w_i w_j$,

is to be a minimum, where the coefficients σ_{ij} denote the variances and covariances of the order statistics of the sample of n.

The means, Ex_i, in the foregoing are the first moments of ranked extremes and have already been tabulated in reference 2. It had been planned to compute the variances and covariances by numerical integration for sample sizes up to n = 10, above which point the computation would have become too costly.

The discovery during the present quarter of a method for expressing the σ_{ij} explicitly in terms of tabulated functions considerably reduces the amount of computation necessary for small samples.

Larger samples can be handled by breaking them into smaller samples, obtaining an estimate of the desired quantity from each one of these samples, and pooling the results. In this manner samples of any size can be handled by the techniques developed for very small samples. This method of subgroups also has the advantage of making possible a control chart procedure whereby internal consistency of the data and stability of operating conditions could be checked. The method is also especially well adapted to the form of estimator being investigated, and would not be applicable to more complicated functions, such as those of Gumbel.

The new method has been rendered still more powerful by a refinement devised by Mr. I. Richard Savage of the Statistical Engineering Laboratory. This refinement not only makes it possible to determine the unbiased estimators for all probability levels P simultaneously, rather than for just a few selected levels such as P = .95, .99, but with very little additional effort yields also the estimators of the two parameters which make it possible to fit an extreme-value distribution to a given set of extreme data. The method also automatically furnishes the efficiency associated with the estimate obtained.

The method has been tried out on a preliminary basis with encouraging results. Unbiased estimators have been found for samples of any size based on splitting the sample into subgroups of n = 2 and 3. If subgroups of 3 are used instead of subgroups of 2, the improvement in efficiency is represented by an increase of 15 percentage points, on the basis of 100 percentage efficiency of the theoretically most efficient estimator. If we proceed from n = 3 to n = 4, 5, etc., further jumps in efficiency are expected to take place, with the result that by the time n = 6 or 7 we should reach 80 or 90 percent efficiency, which is probably adequate for the purpose in view.

C. <u>Proposed further work.</u>

Work thus far points to immediate continuation along the following lines:

(1) Computation of "subgroup" estimators described
4, above for n = 5, 6, 7, say, in order to reach a
practical level of efficiency.

(2) Evaluation of the bias and variance of the present Gumbel-type estimators for comparison with the proposed estimators at least for samples of moderate size. It is proposed that this be accomplished by methods of empirical sampling for n = 10, 20, 30.

(3) Comparison of the results in (1) and (2) with asymptotic theory.

Asymptotic theory for large samples has been discussed in NBS Report 1129, prepared under the NACA project (reference 3). It was there shown how to select as few as three out of a large number n of sample values, which yield unbiased estimators of surprisingly good efficiency. If it turns out that asymptotic theory compares favorably with exact methods even for n as low as 10, then an improvement in procedures might become possible for larger values of n.

III. PLANS FOR PUBLICATION

The following means of writing up the various aspects of the work are envisioned, subject to the usual NACA clearance procedures:

(1) A mathematical paper presenting the derivation of the formula for the evaluation of the double integrals mentioned above, together with related mathematical results.

(2) A technical report to NACA, somewhat in the nature of a manual, which would express the new method in practical terms and would be written for the use of engineering personnel and field workers. This report might also give a general indication of the nature and characteristics of the method.

Julius Lieblein
Statistical Engineering
Laboratory

National Bureau of Standards
Washington, D. C.
14 March 1952

REFERENCES.

1. Press, Harry: The Application of the Statistical Theory of Extreme Values to Gust-Load Problems. NACA TN 1926, 1949.

2. National Bureau of Standards: Table of the first moment of ranked extremes. NBS Report 1167, September 20, 1951.

3. Lieblein, Julius: On certain estimators based on large samples of extremes. NBS Report 1129, August 15, 1951.